Pergamon

Mathl. Comput. Modelling Vol. 22, No. 4-7, pp. 83-101, 1995

0895-7177/95 $9.50 + 0.00

0895-7177(95)00128-X

Estimation of Static Travel Times in a Dynamic Route Guidance System

A. SEN
School of Urban Planning and Policy, University of Illinois at Chicago
Suite 700 S, 1033 W. Van Buren, Chicago, IL 60607-9940, U.S.A.
u08459@uicvm.uic.edu

P. THAKURIAH
National Institute of Statistical Sciences
P.O. Box 14162, Research Triangle Park, NC 27709-4162, U.S.A.
vonuniss.rti.org

Abstract—In an Advanced Traveler Information System where route guidance is provided, a driver chooses a route before he/she actually traverses the links in the route. For such systems, link travel times need to be forecasted. However, information on several thousand links would take a fair amount of time to be conveyed to the driver, and very few drivers would be willing to wait very long to get route information. In the ADVANCE demonstration, to be implemented in suburban Chicago, the in-vehicle unit in each participating vehicle will be provided with the capability of accessing default travel time information, which will offer the vehicle with an autonomous navigation capability. The default estimates will be overwritten by dynamic up-to-the-minute forecasts if such forecasts are different from the default estimates. This paper describes the approach used to compute default travel times estimates.

Keywords—Advanced Travel Information System, Dynamic route guidance, Autonomous route guidance, Link Travel Time Process, Link travel time estimate.

1. INTRODUCTION

A major component of *Intelligent Vehicles and Highway Systems* is *Advanced Traveler Information Systems (ATIS)*. A key feature of ATIS is route guidance which will provide drivers with minimum cost routes to desired destinations. The minimum cost criteria will invariably include travel time. However, route information will have to be supplied to the driver before he or she actually traverses the links comprising the route. Therefore, link travel times must be estimated before the route is computed.

In this paper, we discuss one of the strategies by means of which guidance can be given to drivers participating in an ATIS. The specific system we discuss is a large-scale ATIS with dynamic route guidance capability called Advanced Driver and Vehicle Advisory Navigation Concept (ADVANCE) to be deployed in a suburban area of Chicago. In ADVANCE, up-to-the-minute information on the prevailing traffic conditions will be used in order to help drivers adjust their route choices from the origin of trips to destinations. One of the major purposes of this system is to ensure that underutilized capacities in parts of the road network are used better, while congestion in highly traversed links is mitigated.

In the ADVANCE concept, drivers of participating vehicles will be able to identify a desired destination for their trip. Such vehicles will be equipped with an on board computer called the Mobile Navigation Unit (MNA). The MNA will provide the driver with a route that is desirable according to some criterion, including travel time. Since route advice will be given to the driver

before he or she traverses the links in the route, travel times on the links will need to be forecast. In any Intelligent Vehicle and Highway System (not just ADVANCE) where route advice is given, some form of short-term travel time forecasting would be necessary. A route planning algorithm will reside in the MNA, which will interact with a central computer, called the Traffic Information Center (TIC). TIC has interactive but modular functions and the MNA will construct desirable routes based on travel times that are predicted by the Travel Time Prediction (TTP) module of the TIC.

Vehicles which are equipped with the in-vehicle ATIS components will be called probes. In addition to receiving travel guidance, probes also monitor current traffic conditions. Sensors on board the vehicle and a detailed navigation system will be able to assess the position of the vehicle over short time intervals, in relation to an electronic map of the demonstration area road network, available to the MNA on a CD-ROM. The MNA will be equipped with the ability to broadcast back to the TIC, on an on-line basis using a Radio Frequency (RF) communications system, the time the vehicle takes to traverse each link. Because of RF capacity, not all travel times will be broadcast. Selection methods will be imbedded in the MNA and perhaps activated by the TIC in order to filter real-time travel time reports so that only those travel times which deviate enough from some preestablished travel time norm (to be defined later) for a specific link and time period, will be broadcast.

Route guidance to a participating driver may be available under the above system configuration by means of two strategies: autonomous route guidance and dynamic route guidance. Consider the situation where the MNA receives transmissions of predicted travel times along with other information from the TIC only when the car ignition is turned on. Few drivers would be willing to wait very long to get route guidance after they have entered the car; on the other hand, there are over 10,000 links in the system and the radio frequencies (RF) available have modest capacities. Thus, it may be impossible to broadcast information on all links and have it available for route computation by the MNA. Consequently, default travel time predictions need to be available to the MNA. Moreover, these estimates would also be available in case transmissions are interrupted. These defaults, based on historical and other information, will reside on a compact disk in the MNA and will be called *profiles*. Initial profiles will be constructed using a network equilibrium model off-line for five different time periods.

Autonomous route guidance, therefore, is the strategy whereby, in the absence of TIC-vehicle communication, routes are given to drivers on the basis of default estimates of link travel times which can be accessed by the MNA from a static database available in a CD-ROM. Dynamic route guidance is the strategy in which both the static and the real-time information is made use of in giving guidance.

Some MNAs will also be equipped with memory cards on which travel times on each link traversed by the vehicle, the relevant clock time and other related information will be written. These memory cards would be retrieved and read every few weeks. Based on the information so obtained from

(i) memory cards,
(ii) archived dynamic probe reports obtained at the TIC via the communications system,
(iii) travel time estimates from road detector information, and
(iv) anecdotal information about incidents, constructions, and so on,

profiles will be periodically updated off-line. The updated profiles will then be put in CD-ROMs for drivers to use. They will also be recirculated to the TIC, so that, at any one time, there is compatibility between the profile versions in the car and in the TIC.

This paper will examine the construction of profiles on the basis of which both autonomous guidance and dynamic guidance can be given. Section 2 will present some preliminary information. Section 3 is devoted to methods for profile construction. The case for the need to make updates from one data source to another is made in Section 3.1. Different aspects of the updating procedure are discussed in detail in Sections 4, 7, and 8.

2. PRELIMINARIES

In this section, we present some information on the same theoretical underpinning and practical considerations that motivated the design of the static travel time system. A brief overview of what travel times look like is presented in Section 2.1. This is followed, in Section 2.2, by a description of the data to be used in constructing profiles. The theoretical development of the material that underlies the methods discussed in travel times prediction is given in Section 2.3. In particular, we define what precisely is meant by travel time and the prediction of travel time. Certain implications of using profiles as defaults are given in Section 2.4.
estimation procedure are presented.

2.1. Travel Times: Examples

Suitable direct observations of travel time were not available at the time of development of the static estimation procedure. The examples of travel time that are presented in this section are estimated from flow and occupancy data from road surveillance detectors (see [1] for the method).

Figures 1 through 4 illustrate travel times (computed as mentioned above) for the same location on the inbound Eisenhower expressway in the Chicago metropolitan area leading from the west suburbs to near the Central Business District, for four consecutive Tuesdays. We know of nothing particularly unusual that occurred on or near those links on those days. Thus, if forecasts were based on static estimates alone, these forecasts would be identical for the four days. While during off-peak hours, the travel times illustrated are somewhat similar, during the peaks they are not. Thus, differences between dynamic and static estimates could be noticeable at least on some days—*even in the absence of incidents.* Weekends exhibit a single, long and low peak, as opposed to the typical two-peak weekday pattern. Figures 5 through 8 show travel times estimated at one location on four typical weekend days.

It has been conjectured by some that it might be possible to have the same static estimates for several different links. While this might be possible in some cases, in general, pairs of links seem to require different static estimates, largely due to different land-uses in the surrounding area, especially due to the noticeable differences during peak periods. In order to account for day to day variations, Figure 9 shows average travel times taken over 12 Tuesdays for two contiguous locations on Eisenhower Expressway. Again the differences are obvious during peak hours. It would appear that the land use surrounding a link has enough impact on most links rendering their peak period travel time patterns rather unique.

Day types clearly make a difference in the pattern of travel times. This is illustrated by Figure 10 which shows average travel times on a segment of Eisenhower Expressway for Monday through Thursday (average taken over 12 days—3 for each day of the week), for 6 weekend days (3 Saturdays and 3 Sundays) and 3 Fridays. Friday afternoon peaks are quite different from all others—in fact, in the ADVANCE demonstration, Fridays are a different day-type.

On signalized roadways, traffic signals can have a large effect on travel time reports transmitted from probes. The variances in link travel time due to traffic signals can be quite large, even under undersaturated conditions [2]. This fact, along with the obvious paucity of data for each link (see below) makes the use of probe travel times statistically challenging, to say the least.

2.2. Data

The data needed to make static forecasts will come from several sources:

1. Probe data, via
   (a) Real-time broadcasts of link travel times from the MNA to the TIC which will be retrieved off-line, from data archives.
   (b) Memory card information, which will be more detailed; these will be periodically retrieved (only about 1200 vehicles of the total of about 3000 vehicles will have this capacity).

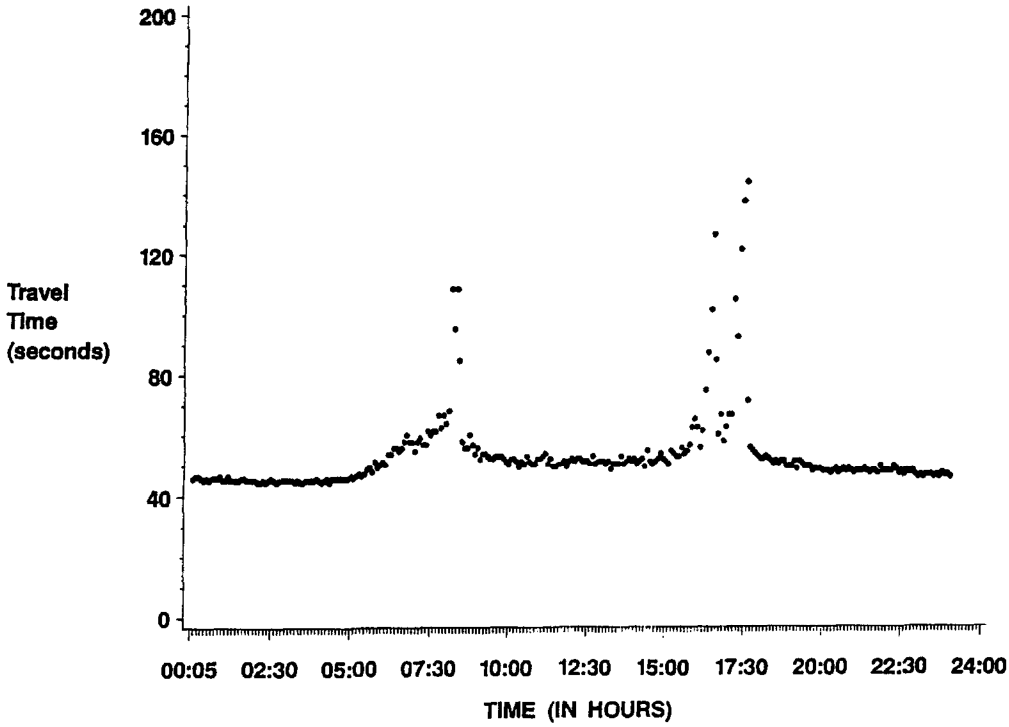

Figure 1. Travel times for Tuesday, October 9, 1990 estimated from detector 3.

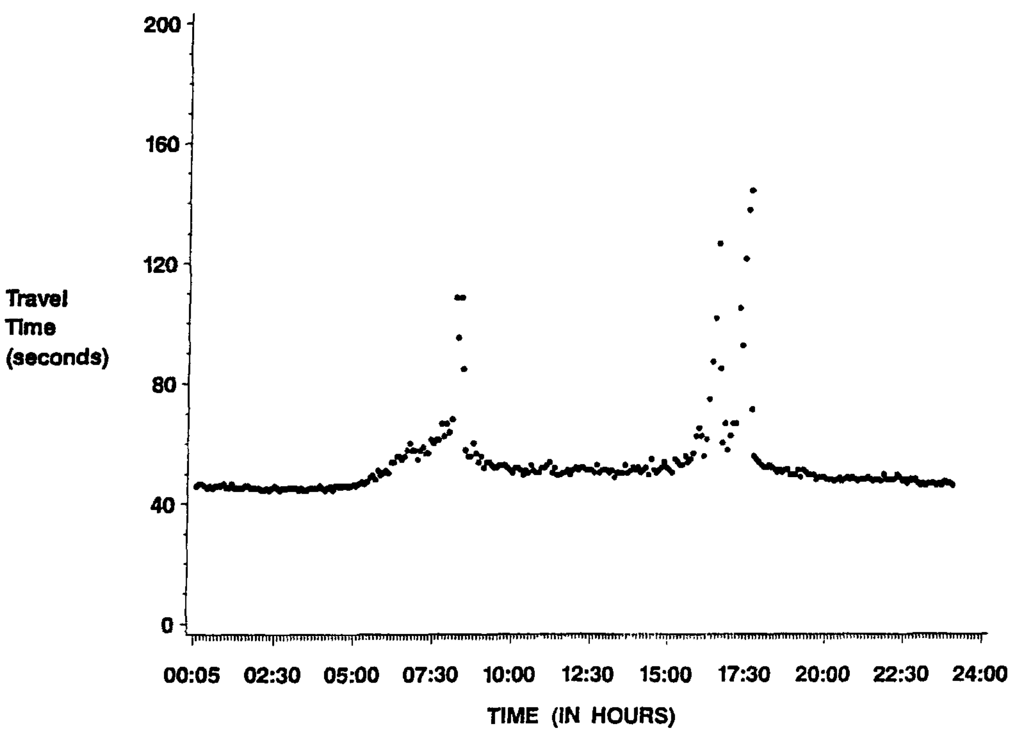

Figure 2. Travel times for Tuesday, October 16, 1990 estimated from detector 3.

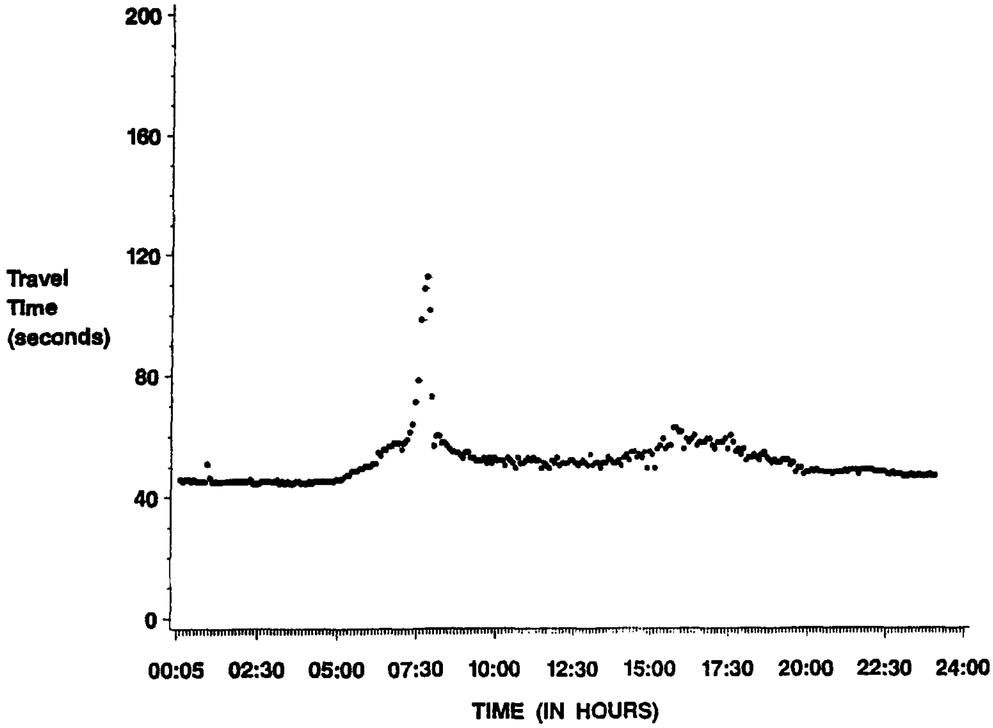

Figure 3. Travel times for Tuesday, October 23, 1990 estimated from detector 3.

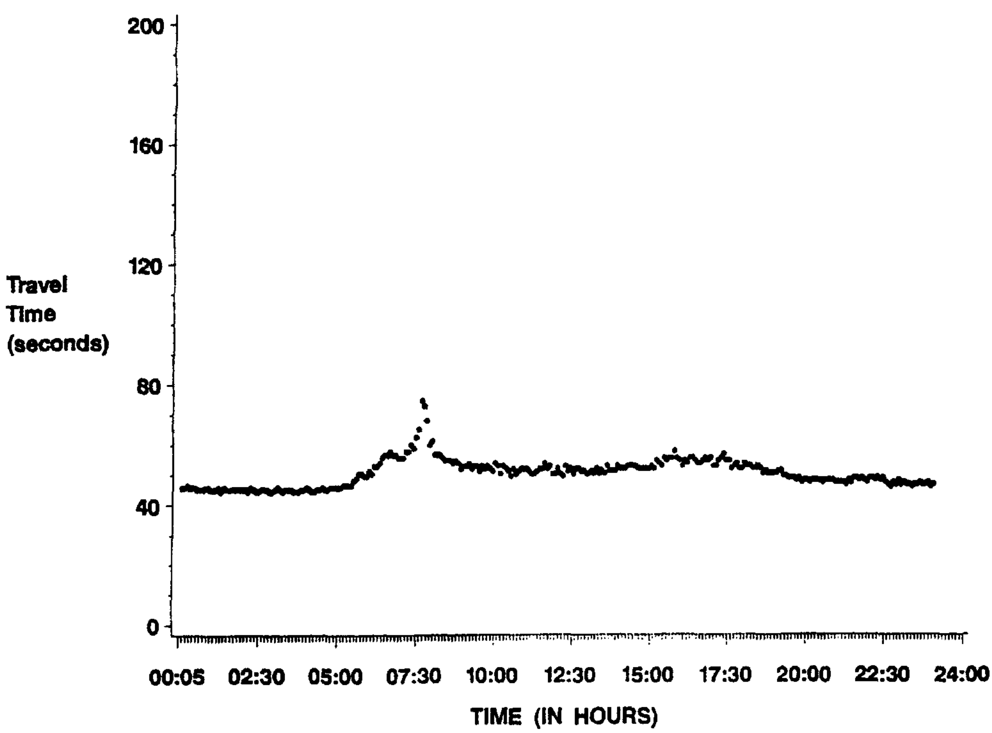

Figure 4. Travel times for Tuesday, October 30, 1990 estimated from detector 3.

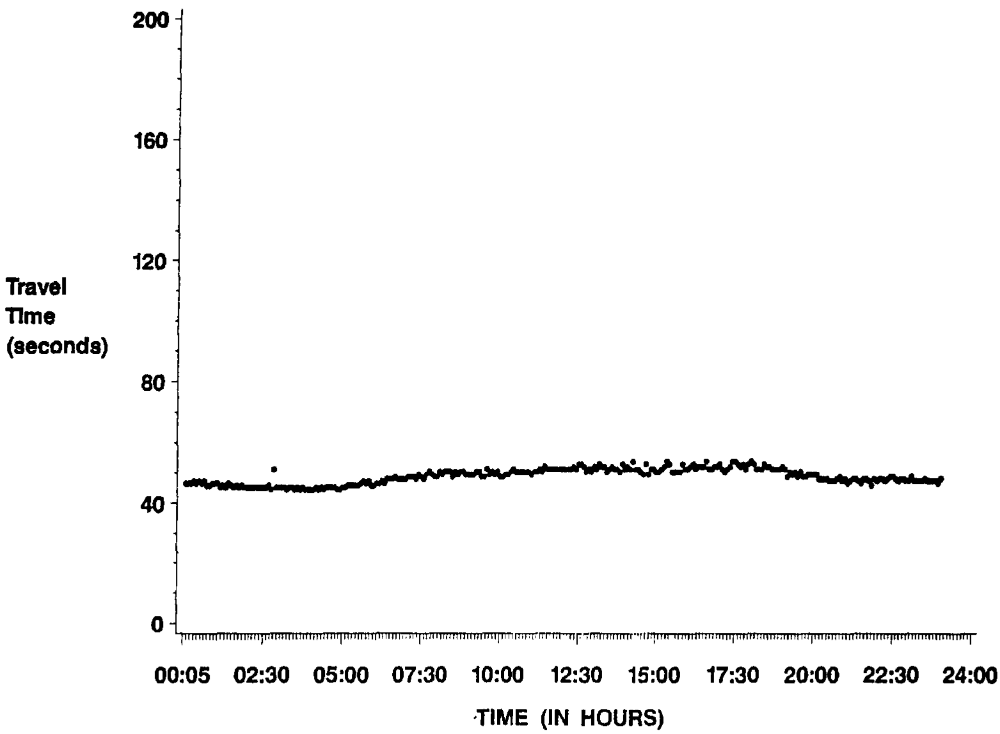

Figure 5. Travel times for a weekend day, Saturday, October 13, 1990 estimated from detector 3.

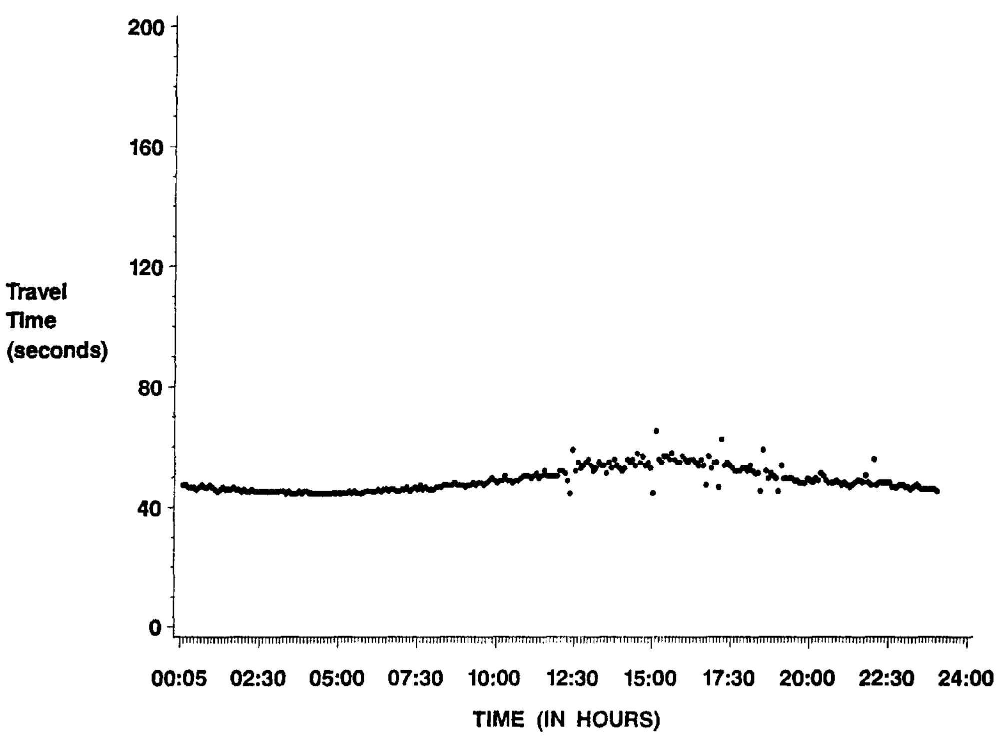

Figure 6. Travel times for a weekend day, Sunday, October 14, 1990 estimated from detector 3.

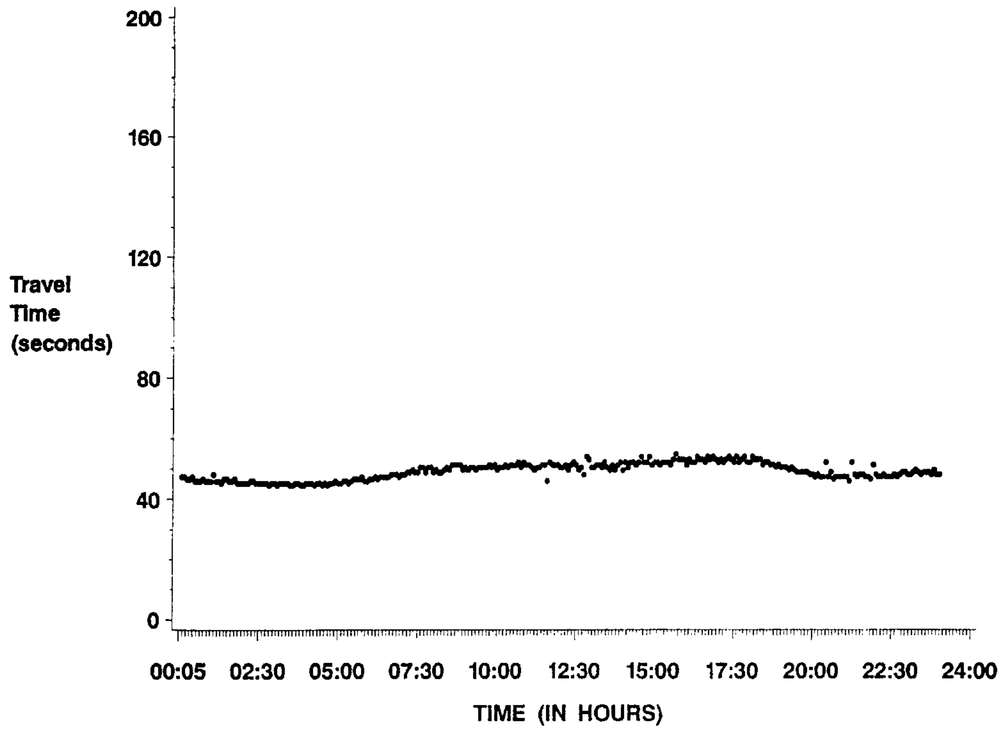

Figure 7. Travel times for a weekend day, Saturday, October 20, 1990 estimated from detector 3.

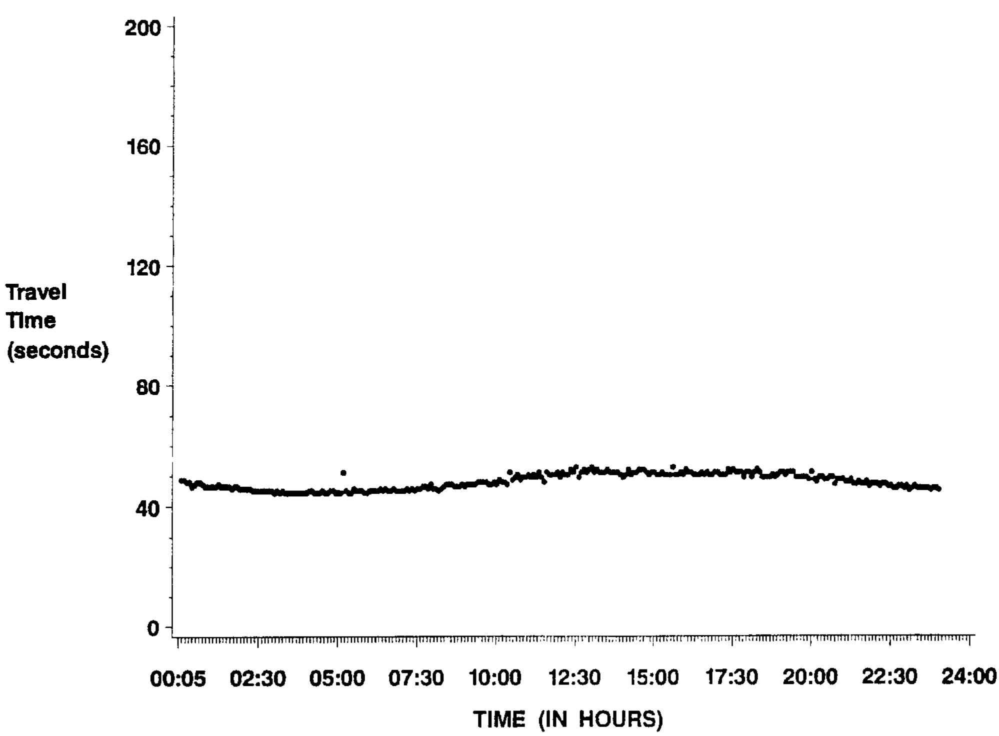

Figure 8. Travel times for a weekend day, Sunday, October 21, 1990 estimated from detector 3.

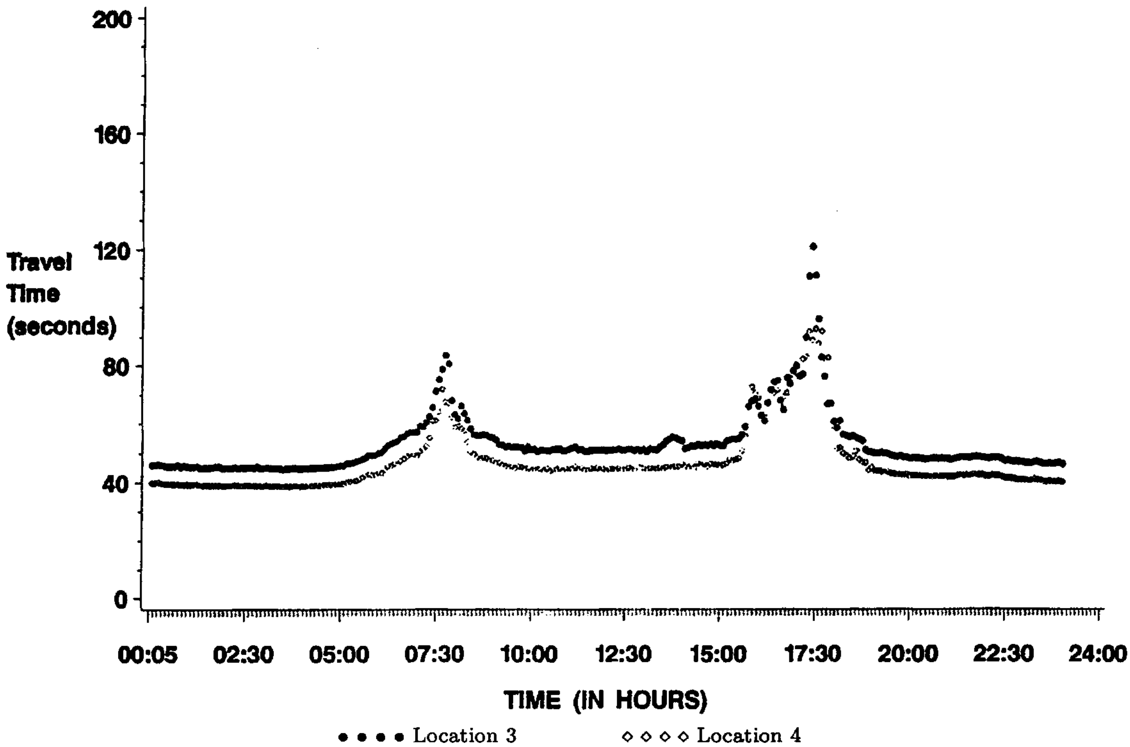

Figure 9. Average travel times for 2 locations taken over 12 Tuesdays. Link travel time profiles. Different location: Same Day_Type (Weekdays without Fridays).

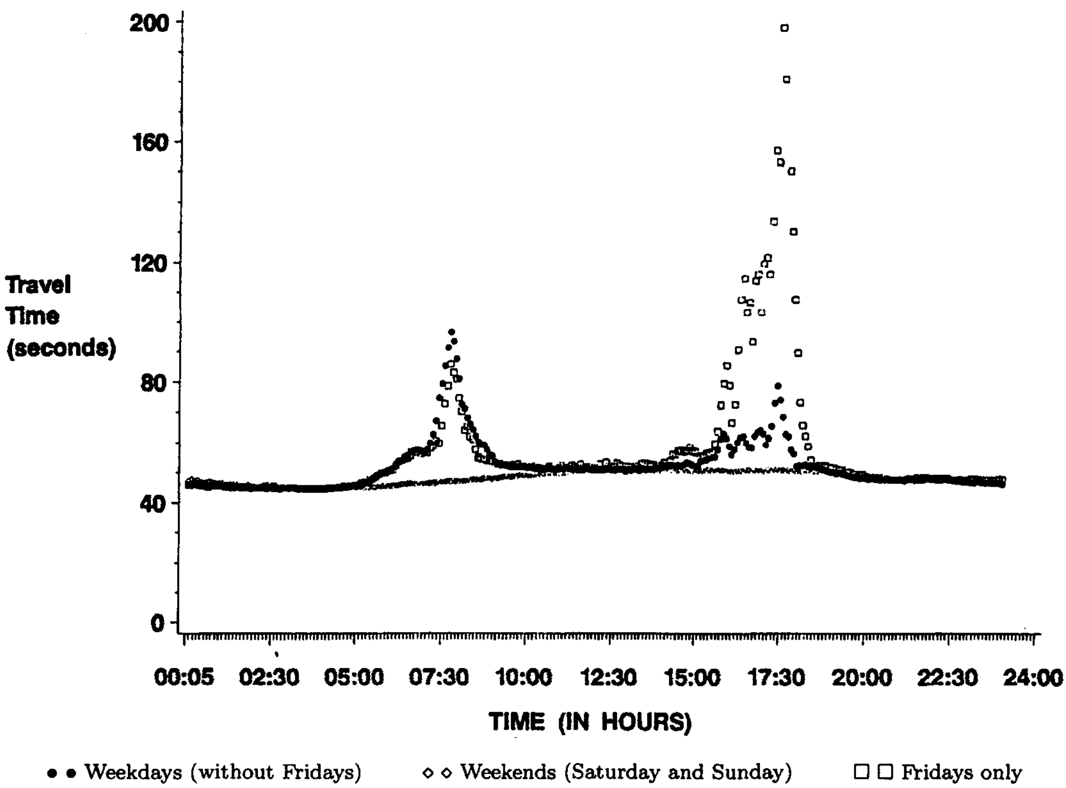

Figure 10. Average travel times for Weekdays without Fridays, Weekends (Saturdays and Sundays) and Fridays only. Link travel time profiles. Same location (Location 3): Three different Day_Types.

2. Volume and occupancy data obtained from road detectors, from which estimates of travel times will be made. Such data will be available from a limited number of links in the ADVANCE study area.
3. Anecdotal data, principally those about incidents, which will be used to identify observations to be deleted from consideration.

Probe data are the principal source of information on travel times. However, the biggest constraint on this source of information is the sparseness of the data which is of great importance in the design of statistical procedures. The demonstration area road network is large and the number of roving probes limited. It is, therefore, important to ask how much data from equipped vehicles would be available under different conditions; i.e., how many appropriately equipped vehicles would traverse each link under different conditions. Hicks *et al.* [3] have estimated that if 5000 vehicles are a part of the demonstration and are in the target area during rush hour, more than 40 percent of the links will not have any equipped vehicles on them during a typical five-minute period and more than 20 percent will not be traversed by a probe in 15 minutes. Moreover, except for freeways and major arterials, very few links will have more than 2 such vehicles on them during a ten minute period. For freeways and arterials, the numbers are higher but not excessively so. Given the rather large variances of the distribution of travel times at any given moment, the amount of data that would be transmitted from every link is clearly small over short intervals of time, even during peak hours, when deployment levels (number of vehicles with ATIS components) are low. During off-peak hours, there will be even fewer probes on each link. For left turn links, particularly for weekend days, data will be even more sparse.

An added issue is that some dynamic probe reports may be "unreliable" due to equipment drawbacks or traveler actions (for instance, a driver may stop the vehicle for a short time to perform some activity). These reports will be discarded or somehow filtered by another component in the TIC, called Data Fusion.

2.3. The Link Travel Time Process

It is important at this stage to describe the meaning of link travel time. While the concept is loosely understood, the precision is critical for any statistical analysis. For the present purpose, the physical link is defined as a stretch of roadway after the downstream intersection and including the upstream turning movement (through movement and one or more left and right turn).

2.3.1. Definition of the Link Travel Time Process

For a given link ℓ, a specific car with a specific driver, entering the link at time t, will have a certain travel time. There is also a probability that this car-driver combination will enter the link at time t. Therefore, the marginal expectation of travel time over all car-driver combinations is well defined. Call it the travel time $T_{\ell,t}$ on link ℓ at time t of entering the link.

For each t, $T_{\ell,t}$ is a random variable. Intuitively, it seems obvious that it has a finite mean and a reasonably well-behaved distribution. $T_{\ell,t}$ for all t defines a stochastic process which we shall call a Link Travel Time Process or LTTP. Figure 11 shows a schematic diagram of such a process. The solid curve shows the mean link travel time $\mathrm{E}[T_{\ell,t}]$ for any (clock) time t over a 15-hour period (for a specific day). For any fixed value t, $T_{\ell,t}$ is a random variable with a distribution. The density function of $T_{\ell,t}$ at time t_0 is shown as a dashed curve. One might wish to view the density curve as being drawn in a plane orthogonal to the plane of the paper and also recall that corresponding to every point on the solid curve, there is such a density.

It is important to stress that for any time t, the random variable $T_\ell(t)$ exists. However, we may be able to obtain observations only for specific values of t, corresponding to the existence of probes on link ℓ.

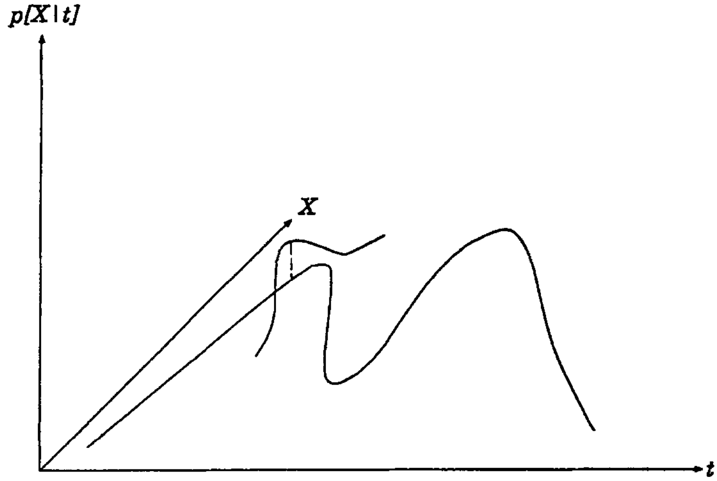

Figure 11. Illustration of a Link Travel Time Process.

It should also be noted that, for the present, we do not assume that the LTTP for two days need be the same; i.e., t is clock time in the sense of a calendar and clock (2 p.m. on March 3 and March 4 are two distinct values of t).

2.3.2. Description of the problem

To compute a shortest route between the origin of a trip and the destination, the MNA needs, for each time t and for each link ℓ, a single number which is the predicted travel time. Thus, we need to estimate the expectation $\mathrm{E}[T_{\ell,t}]$ of $T_{\ell,t}$ for future values of time t. Since $T_{\ell,t}$ varies widely depending on conditions on the link, we need to estimate a conditional expectation, $\mathrm{E}[T_{\ell,t} \mid \mathbf{c}]$, where $\mathbf{c}$ is a set of conditions that need to be determined. Estimation of a conditional expectation is a general form of regression. In this case, the components of the vector $\mathbf{c}$ are the 'independent variables.'

Since arbitrary functions are impossible to store on a hard disk or transmit over radio, approximations are needed even for estimates of $\mathrm{E}[T_{\ell,t} \mid \mathbf{c}]$. The type of approximation chosen for the ADVANCE project consist of step functions. That is, we shall assume that clock time is divided into intervals and over each interval $\mathrm{E}[T_{\ell,t} \mid \mathbf{c}]$ is constant.

Largely because of radio frequency constraints, it is convenient to partition $\mathbf{c}$, which is a vector containing measurements on the variables affecting travel time, into three parts: $\mathbf{c}_1$, $\mathbf{c}_2$, and $\mathbf{c}_3$. Part $\mathbf{c}_1$ will consist of values of variables that will be known well in advance, such as day of week, time of day, lane blockages due to construction, special events, or variables whose effects may be known in advance, such as suitably coded weather conditions. Another variable that needs to be considered will be mentioned later. Part $\mathbf{c}_2$ could consist of values of variables which will not be known until shortly before time t. For instance, travel time on the link just before time t, travel times on upstream links or downstream links (which we do not propose to use but are mentioned here in order to maintain the generality of the procedure) and incidents could be in Part $\mathbf{c}_2$. Part $\mathbf{c}_3$ will consist of values of variables that will be too difficult to measure (for example, the relation of $T_{\ell,t}$ to the traffic signal cycle) or variables whose effects are considered negligible in the estimation problem such as the effect of vehicle dynamics or driving styles on $T_{\ell,t}$.

Write

$$\begin{aligned}
T_{\ell,t} &= \mathrm{E}[T_{\ell,t} \mid \mathbf{c}_1\mathbf{c}_2\mathbf{c}_3] + \eta_{\ell,t} \\
&= \mathrm{E}[T_{\ell,t} \mid \mathbf{c}_1] + \big(\mathrm{E}[T_{\ell,t} \mid \mathbf{c}_1\mathbf{c}_2] - \mathrm{E}[T_{\ell,t} \mid \mathbf{c}_1]\big) + \big(\mathrm{E}[T_{\ell,t} \mid \mathbf{c}_1\mathbf{c}_2\mathbf{c}_3] - \mathrm{E}[T_{\ell,t} \mid \mathbf{c}_1\mathbf{c}_2]\big) + \eta_{\ell,t} \qquad (1) \\
&= \mathrm{E}[T_{\ell,t} \mid \mathbf{c}_1] + A_{\ell,t} + B_{\ell,t} + \eta_{\ell,t}
\end{aligned}$$

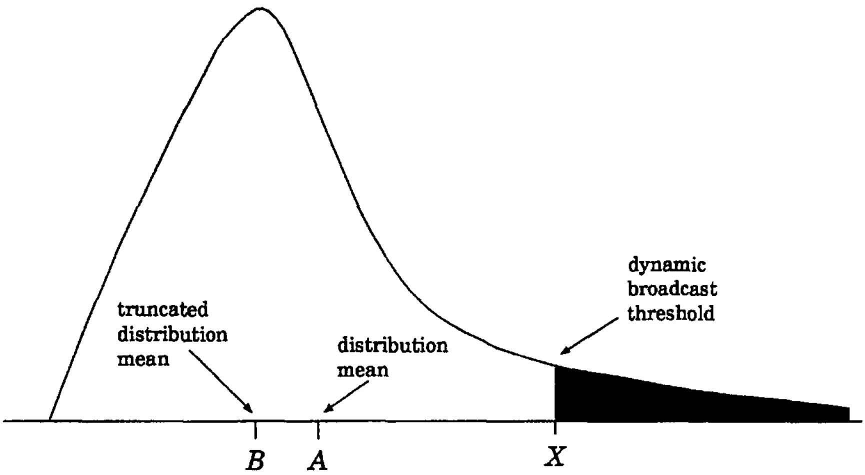

Figure 12. Difference between biased and unbiased estimates of expected link travel times

where

$$A_{\ell,t} = \left(\mathrm{E}[T_{\ell,t} \mid c_1 c_2] - \mathrm{E}[T_{\ell,t} \mid c_1]\right),$$
$$B_{\ell,t} = \left(\mathrm{E}[T_{\ell,t} \mid c_1 c_2 c_3] - \mathrm{E}[T_{\ell,t} \mid c_1 c_2]\right)$$

and $\eta_{\ell,t}$ is an error term. Since Part c_3 is too difficult to measure, we will ignore $B_{\ell,t}$ and assume it to be absorbed into the error.

We shall call an estimate of $\mathrm{E}[T_{\ell,t} \mid c_1]$ a static estimate and an estimate of $\mathrm{E}[T_{\ell,t} \mid c_1 c_2]$ a dynamic forecast. Clearly, in either case, these estimates are being constructed at some time before t; i.e., t is a time in the future. As mentioned in Section 1, because of radio frequency capacity, we need to economize on the number of messages; therefore, we shall consider estimates of $\mathrm{E}[T_{\ell,t} \mid c_1]$ as a default, providing dynamic forecasts only when needed—ideally when $A_{\ell,t}$ is large enough. In the next section, we show that an unbiased estimate of $\mathrm{E}[T_{\ell,t} \mid c_1]$ may not be the best estimate to use as default. Moreover, the conditions under which dynamic forecasts are broadcast would determine how biased the static estimates need to be. However, there might also be occasions where unbiased estimates might be needed. Thus, we might need several estimates of $\mathrm{E}[T_{\ell,t} \mid c_1]$. For future reference, an unbiased estimate of $\mathrm{E}[T_{\ell,t} \mid c_1]$ will be called a static estimate of the Link Travel Time Process, while a biased estimate which is more suited to use as default is named a link travel time profile.

2.4. Profiles as Defaults

While profiles have been defined as estimates of expected values of travel times, their use might require that they be biased estimates. The problem is best illustrated with an example. Figure 12 illustrates a distribution of travel times for a link at some specified time, which is skewed to the right. The mean of the distribution is given by the point A. Suppose a dynamic forecast is given whenever travel times exceed X. Then observations exceeding X should not be considered in obtaining the profile. The expectation of travel times less than X is the expectation of the truncated distribution (shown unshaded), and is equal to B.

Thus, profiles should be B. Otherwise, in the long run, wrong average travel times would occur. To see how this occurs and what the solution to the problem might be, consider a highly simplified and indeed unrealistic problem. Assume (to keep this illustration computationally simple) that the distribution of link travel times is discrete and the values 1, 2, 3, 4, and 5 minutes each have a 20 percent chance of occurring. The mean travel time is readily seen to be 3 minutes. Now suppose that whenever the travel time is actually 5 minutes, we are told that it is 5 minutes. Thus, 80 percent of the time we would use the mean 3 minutes as the default estimate, and 20 percent of the time, we use the actual value of 5 minutes. We would end up with an average value of $0.8 \times 3 + 0.2 \times 5 = 3.4$ minutes.

Let the correct default estimate be x. In the long run, we want our estimate to approximate the mean. Thus, $0.8x + 0.2 \times 5$ needs to be 3 minutes. That is, $0.8x$ needs to be 2, or x needs to be 2.5, which is the average of 1, 2, 3, and 4. The principle here is that if we use a mean to stand in for a set of values, the mean needs to be that of only the values it is standing in for.

The general formula for constructing default estimates of travel times may be presented as follows. Let X be the random variable, link travel time, and let its density function be $p[X]$. Further, for each value X of X, let $\pi[X]$ be the probability that the true value X would not be known to the MNA. Then the correct default estimate (or profile) y is

$$y = \frac{\int_{-\infty}^{\infty} X p[X]\, \pi[X]}{\int_{-\infty}^{\infty} p[X]\, \pi[X]}. \tag{2}$$

A proof follows: Let Y be the default or profile value. Then, for a given value of X, there is a probability $1 - \pi[X]$ that we would know it, and hence, the expected value that would be used by the MNA would be

$$\pi[X]Y + (1 - \pi[X])X.$$

Taking expectations over all X and equating the result to the expectation of X, we have

$$\int_{-\infty}^{\infty} [\pi[X]Y + (1 - \pi[X])X] P[X] = \int_{-\infty}^{\infty} X P[X].$$

Hence,

$$Y \int_{-\infty}^{\infty} \pi[X]\, P[X] = \int_{-\infty}^{\infty} X\, P[X] - \int_{-\infty}^{\infty} X\, P[X] + \int_{-\infty}^{\infty} \pi[X]\, P[X] X$$

and (2) follows on solving for Y.

3. STATIC FORECASTING

As mentioned in Section 2, static estimates and link profiles are estimates of what we have called static times, which are link travel times under easily predicted conditions (e.g., time of day). In this section, we examine the construction and updating of static forecasts and also of profiles (see Section 2.4).

The compact-disks in probes will initially be supplied with link travel times from a network flow model (NFM). The network flow model is essentially a network equilibrium model, but with realistic delay functions incorporated. These times will be updated with estimates of link travel time based on actual observations as and when probe data become available.

It should be noted that the NFM will only be run for weekdays. For weekends, initial values will need to be constructed differently. We will refer to the transition from NFM link travel time figures to static estimates of link travel time based on NFM times *and* probe and other data mentioned in Section 2.2 as the *initial update* of a link travel time estimate (see Figure 13). Further, these initial update travel time estimates will be modified as more probe and other data are gathered. This is expected to be a continuous process which will last throughout the life of the demonstration. The second and subsequent updates made to each link travel time estimate will be referred to as *modification updates*.

3.1. Need for Updates

This section outlines why neither NFM estimates nor estimates based on probe data alone would be entirely satisfactory for static forecasting. The two information sources are, in fact, better combined. Several potential deficiencies *may be* associated with travel time estimates from the NFM. These include:

1. Very little empirical verification of such times has occurred. Therefore, it is unknown how closely such estimates match actual travel times, especially when the network becomes congested as in peak hours.

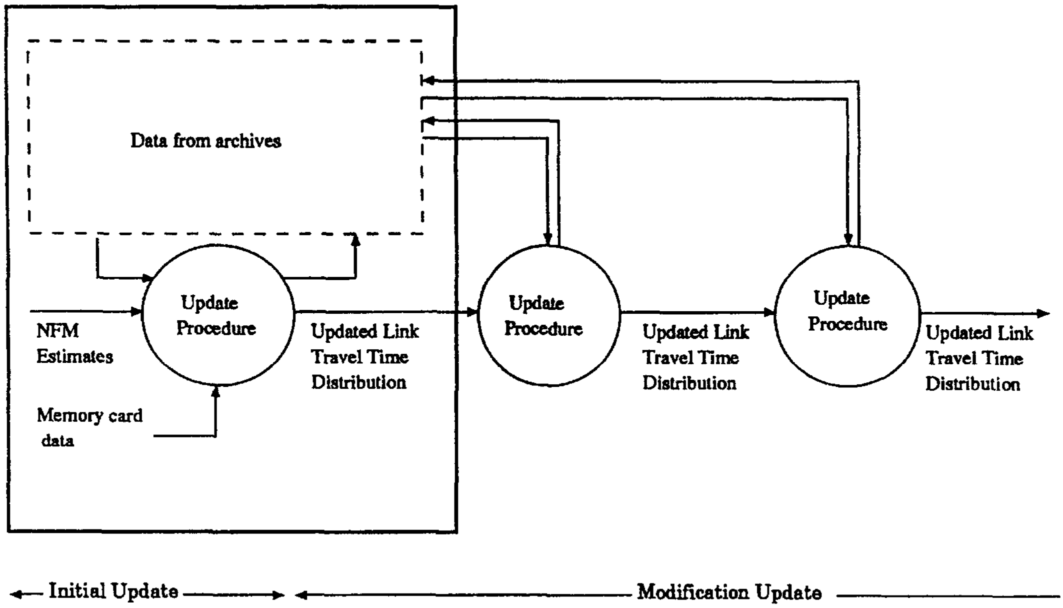

Figure 13. Updating process.

2. The present plan is to run a network equilibrium model for five different periods during weekdays. As a result, we will get travel times as a function of clock time and the function will be a step function consisting of a limited number of travel time levels.
3. There is no provision within operational network equilibrium models as they currently exist to consider unusual circumstances such as weather.

While, because of the concerns mentioned above, one might be tempted to suggest that travel times from network flow models be replaced by estimates based solely on times reported by probes, such probe data have deficiencies too:

1. Travel time reports from probes will have very large variances. This problem will be compounded by measurement and reporting errors made by the MNA.
2. Initially, such data will be very sparse. Even if data are accumulated over, say, a 3-month period, very few observations would be available for certain 15-minute intervals for certain links—for instance, those with turning movements. This would be particularly true for Saturdays and Sundays.
3. Observations for unusual days such as days with snow or rainy conditions would be even more scarce.

For these reasons, a combination of NFM estimates and estimates based on probe data is proposed.

4. UPDATING PROCEDURE

A well-known approach in statistics which is to combine an initial estimate with some data to come up with a improved estimate is called Bayes' method. To apply Bayes' method, we need a distribution of the initial estimate (called a prior distribution) and some data on the phenomenon one is estimating. The method then yields a revised distribution for the estimate we are seeking. This distribution is called a posterior distribution.

After the initial update is made using Bayes' procedure, we would have a distribution for the estimate. This distribution would be input into the next modification update procedure (along with additional travel time data) and will result in an improved estimate and an improved distribution for it. And this process can continue.

Bayes' Theorem is quite straightforward: if $p(\theta \mid x)$ is the conditional probability density of θ, given x, and $p(\theta)$ is an unconditional density, then

$$p(\theta \mid x)\, p(x) = p(\theta) p(x \mid \theta). \tag{3}$$

The proof is equally straightforward; both sides of the above equation equal the joint probability density $p(\theta, x)$. In the present context, θ is an estimate (from the NFM) of mean travel time for a particular time of day and a particular day type, and x is the vector of probe reported travel time. $p(\theta \mid x)$ is the updated distribution of θ, $p(\theta)$ is its prior distribution and $p(x \mid \theta)$ is the distribution of x given θ. $p(x)$ can be computed from

$$p(x) = \int p(\theta) p(x \mid \theta)\, d\theta.$$

While no particular difficulty exists in using the above formula, life is considerably simplified if we are able to assume that all distributions are normal. Then, if $\bar{x}$ is the mean of probe reports, θ_0 is the initial value of θ, σ_0^2 is the initial variance of θ and $\sigma_{\bar{x}}^2$ is the variance of $\bar{x}$, then the posterior distribution of θ is normal with mean

$$\theta_1 = w_0 \theta_0 + w_1 \bar{x}, \tag{4}$$

and variance σ_1^2 given by

$$\sigma_1^{-2} = \sigma_0^{-2} + \sigma_{\bar{x}}^{-2}, \tag{5}$$

with $w_0 = \sigma_1^2/\sigma_0^2$ and $w_1 = \sigma_1^2/\sigma_{\bar{x}}^2$. Equation (5) is often written as

$$\text{Posterior Precision} = \text{Prior Precision} + \text{Datum Precision}.$$

These formulæ have been derived in [4]. The main point is that there are fairly straightforward formulæ for updating link travel time estimates, if

1. normality can be assumed, and
2. all the terms involved are known.

We examine each of these premises next. Since $\bar{x}$ is a mean, as long as there are enough probe observations, approximate normality can be assumed. The NFM is not a statistical model; however, since the attempt is to construct a mean-like estimate, it is not too much of a reach to assume normality as long as the estimate is reasonably mean-like—it needs to be at least close to unbiased.

θ_0 will be supplied by the NFM, and $\bar{x}$ can be easily computed from probe reports, as can an estimate of the variance $\sigma_{\bar{x}}^2$ of $\bar{x}$. Thus, all we need is σ_1^2.

Two further points of interest are:

1. If the mean travel time for a particular time of day actually remains unchanged, eventually, with enough probe reports, we will get estimates of the mean travel time which are very close to the true value. On the other hand, if the mean shifts, we need to revise our procedure as we discuss below.
2. If the quality of the initial estimate is poor, or nonexistent, by setting $\sigma_0^2 \to \infty$, the estimate of θ_1 given in (3) becomes $\bar{x}$ as should be expected. This is useful for weekend days when NFM estimates are not available or when, because of shifts in the mean, no prior information is available.

5. ADJUSTMENT OF NFM ESTIMATES

The NFM will supply θ_0, which is the "initial estimate." This will essentially be one number for every link (by turning movement). However, in order to update from NFM numbers, we need the corresponding σ_0^2 or the variances of the NFM estimates. Note also that (from (4)) although

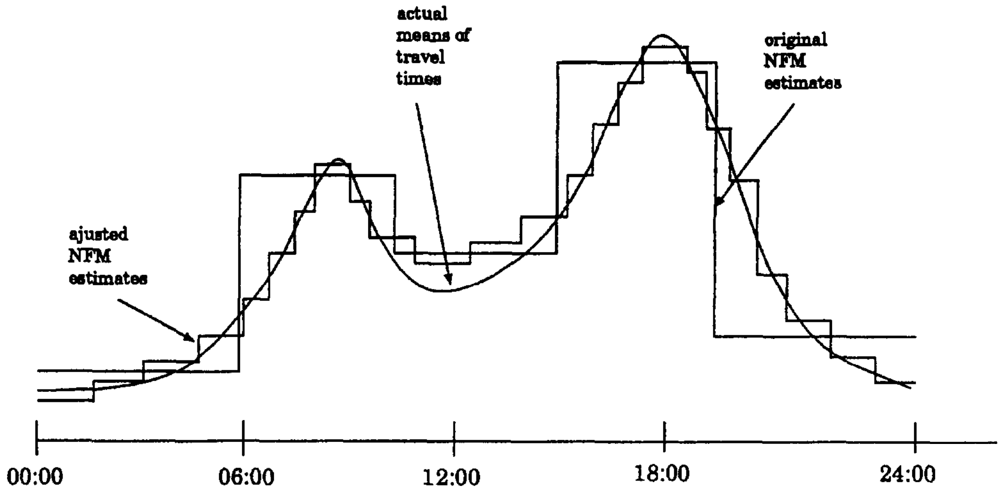

Figure 14. Partial smoothing of step functions.

a reasonable estimate of the variance σ_0^2 of each NFM is necessary for the procedure to work, it does not have to be too precise, since its role is mainly to tell us how much emphasis to give the NFM estimates.

Two possibilities exist in order to obtain estimates of the variances of NFM estimates.

1. In one, we would cast the problem in a regression framework. An illustrative example, using NFM information is

   NFM left-turn travel time

   $$= F(\text{NFM through travel time, functional class, other relevant factors}).$$

   The above will yield the standard error of prediction, which will be an estimate of σ_0^2 for the NFM left turn link travel time estimate. Of course, we would also need to replace the NFM values with the corresponding predicted values.

2. Alternatively, we could simply calculate averages of squared deviations of observations from NFM estimates (mean square error).

Both approaches will be tried out and one selected. However, the regression approach has an additional advantage. Before the first update, a very limited amount of probe data will become available. This information will be used to smooth the step-function aspect of the NFM estimates. A regression approach, such as the one above could enable scaling of the NFM estimates so that we get adjusted NFM estimates. Construction of a different F for each time of day would accomplish partial smoothing of the NFM "levels" as shown in Figure 14.

Temporal consistency becomes an issue as well when a procedure such as the above is used. Let the NFM estimate for a through link be 1 minute and that of its mate left turn link be 3 minutes. Suppose from the limited probe data it is estimated that the through link travel time is actually 3 minutes and we have no probe data at all for the corresponding left turn link. The estimate for the left turn link would remain 3 minutes. But since the estimated difference between the left turn movement travel time and through movement travel time is 2 minutes, the adjusted left turn travel time now becomes 5 minutes. Transferring of estimates from similar links may also be done for links on which there is no data at all before the initial update.

6. OPERATIONALIZATION OF PROCEDURES

In the initial updating procedure, we would have NFM estimates for weekdays. The NFM would yield estimates of link travel times over five periods for weekdays, as shown in Figure 14. If steps are carried out as described in the last section, we would be reasonably assured of their

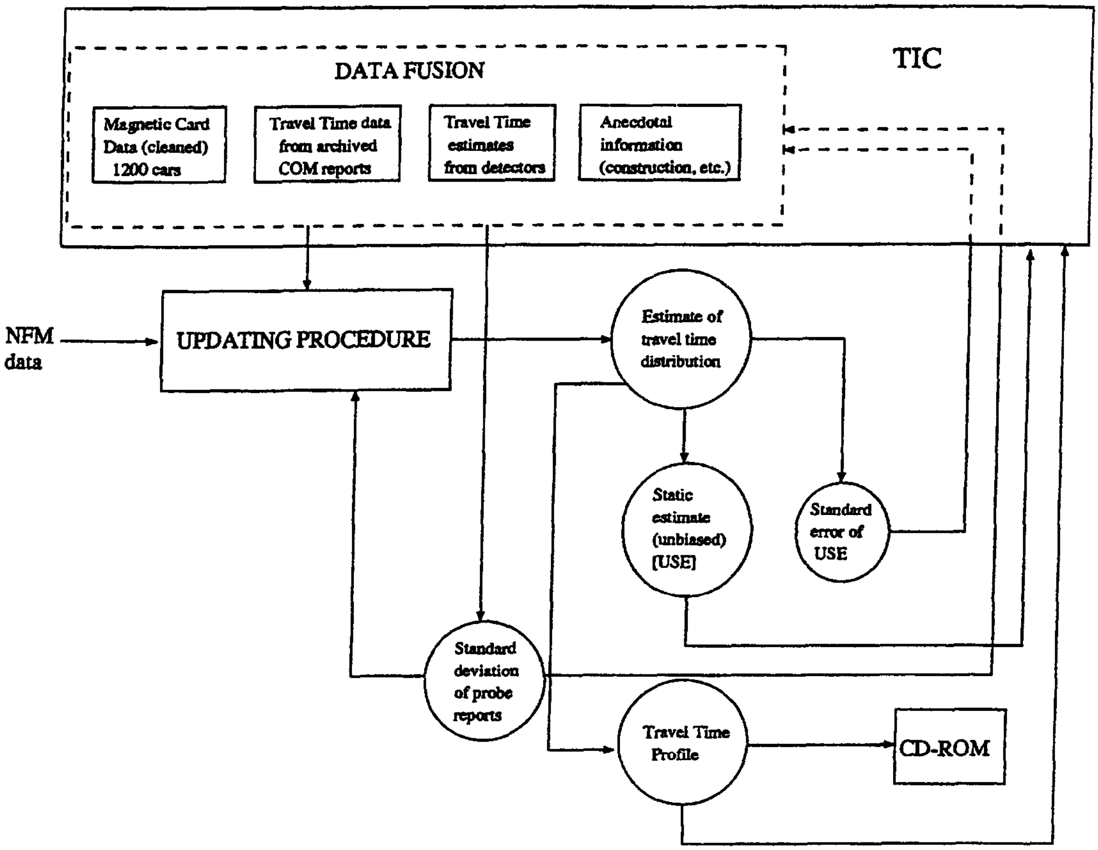

Figure 15. Procedure for obtaining static estimates and profiles.

unbiasedness and also have their variances. In Section 2.2, the data that would be used has been described. From these data, the arithmetic mean of travel times as well as an estimate of its variance can be computed for every time interval. These are the inputs for Bayes' procedure. The output is the posterior distribution, from which a static estimate can be computed for every time interval, as well as the variance of this estimate. The adjusted estimates of travel time are also shown in Figure 14.

Given the posterior distribution as well as the empirical distribution of vehicle of travel times, if a strategy is known by which dynamic forecasts will be selected for broadcast, profiles can be created. This posterior distribution will also be the prior for the next update. The profiles will be sent to the MNA, the static estimates and their standard errors would be reinput into TIC for future use.

As mentioned before, no NFM estimates would be available for weekends and holidays. In those cases, the initial estimate of travel times will simply be the mean of vehicle travel times. As we discussed in Section 2.4, profiles could be different from static estimates. Profiles are the averages only of travel times that are within a typical range—i.e., a range outside of which we would tend to broadcast dynamic estimates to replace profile values. Formula (2) gives a method of computation of such profiles. Note that the density $p(x)$ in that formula refers to a distribution of travel times, not a distribution of estimates.

We conclude this section by describing the way in which (2) can be made operational. First, since distributions of travel times need to be estimated empirically, we would start with a histogram of travel times (for a given clock-time interval). Those bars of the histogram which correspond to travel times for which dynamic estimates would be broadcast would be excluded. This would be for the whole bar if dynamic broadcasts would always be made for the corresponding travel times. If, say, broadcasts would be made for α percent of the time then α percent of

the bar would be removed. Then we would construct an average of travel times which is weighted by the bars (or portions of bars) that have not been removed.

Obviously, to carry out this computation we would need:

1. a histogram of travel times for a given link and given time interval, and
2. information on when dynamic forecasts would be broadcast.

Fortunately, after initial deployment of equipped vehicles, no dynamic broadcasts will be made. During this time, profiles would be the same as static estimates. Also, during this time, we would gather data so as to construct the histogram mentioned above.

7. MODIFICATION UPDATES

Modification updates will be constructed in much the same way, with MNA estimates being replaced by estimates from the previous update. The variance of these estimates would be known from the update.

There is, however, one critical difference between the initial updates and later updates: in the above discussion, we have assumed that the distribution of travel time on a specific link at a specific time is the same—only our uncertainty came from a lack of knowledge of it. However, when data are gathered over a long period, distributions could change owing to a number of factors including changes in land use (for instance, a new building could generate additional traffic or lane closures due to long-term construction). That is, newly gathered data could reflect a different underlying situation than the estimates constructed at a previous update. Therefore, one needs to constantly check if the travel time distribution has changed. Well-known tests exist in the statistics and quality control literature (see [5–9]).

The test that will be used is a likelihood ratio test [7]. Let $x_1, x_2, \ldots, x_N$ be N probe reports for link ℓ over a time period between two consecutive updates. Let the means be $\mu_1 = \mathrm{E}[x_1]$, $\mu_2 = \mathrm{E}[x_2]$ and so on, with unknown common variance σ^2. The test is used to assess if at some unknown point r $(1 \leq r \leq N)$, in which μ_i changes to μ_i^*. The problem then is to test the hypothesis

$$H : \mu_1 = \mu_2 = \cdots = \mu$$

against the two-sided alternative

$$A : \mu = \mu_1 = \cdots = \mu_r \neq \mu_{r+1} = \cdots = \mu_N = \mu + \delta,$$

where r, the change point and δ, which is the shift, are unknown. A statistic for the above problem based on a maximum likelihood estimate of r is

$$s = \sup_{1 \leq r \leq N} \left\{ \frac{(N-r)\bar{x}_{N-r}^2}{(N-1)^{-1}\left(\sum_{i=1}^{r} x_i^2 + \sum_{i=r+1}^{N} (x_i - \bar{x}_{N-r})^2\right)} \right\}.$$

Before the updating procedure begins, the test would be run to see if there has been a change. Figure 16 illustrates the steps involved in operationalizing the test. Three outcomes are possible:

1. The test indicates that conditions are more or less the same. The usual updating procedure will be carried out.
2. There is some uncertainty. The usual updating procedure will be carried out, but now the link will be flagged so that it is scrutinized carefully at the next updating exercise.
3. There is a strong chance that a fundamental change has occurred. Then new estimates will be constructed according to the usual procedure and the link will also be marked for scrutiny at the next updating effort. If at the next update, the data indicates that a change has occurred, the new estimate, based only on observations after r will form the basis for the update. If not, we would return to the estimates we were using before suspecting the change.

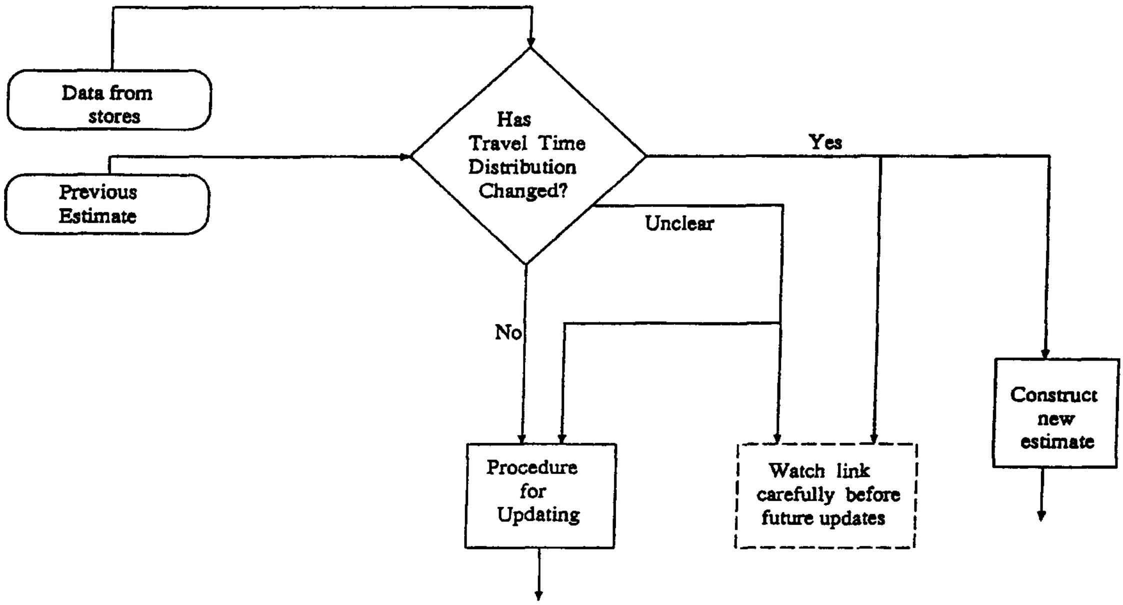

Figure 16. Modification updates.

8. UNUSUAL CIRCUMSTANCES

As discussed in the last section, for certain links, there might be very little data for certain time periods. In fact, even if we accumulate data over, say, three months, there might be very little probe generated data for certain links and at certain times. This would be particularly true of weekend days and even more so for other holidays. Moreover, if we need data for different levels of inclemency of weather, then observations on particular links, day types, times of day *and* different levels of inclemency would be most scarce.

Therefore, the effects of a certain level of inclemency of weather might have to be determined on the basis of a few links and extrapolated to other links. One method for doing this would be to use a large linear (regression) model. The dependent variable would be link travel time (actually, most likely a transformation of link travel time—probably a log transformation) and independent variables would be indicator variables (which are variables taking only two values—often 0 and 1). Each indicator variable would represent a different level of a factor. For example, one indicator variable could represent 'light snow'; it would take the value 1 when snow is light and the value 0 otherwise. The factors could be link type, day type, time of day, weather condition and so on. (For more information on the representation of different levels of several factors by indicator variables, see [10, Chapter 4].) In order to prevent the exercise from becoming too large, some grouping might be desirable. However, notice that because the effect of, say, snow would be quite different at different times of day, there will be need for interaction terms.

The result from such a regression exercise would be formulæ which would enable adjustment for different conditions. For example, if a log transformation were used, we could get an estimate that says that all travel times need to be increased by a certain percentage if there is, say, light snow.

Clearly, such a regression exercise needs to be carried out carefully, in accordance with well-known procedural standards. Several problems can be anticipated even at this stage and these problems would have to addressed. Included among these problems are correlated errors and heteroscedasticity. In addition, there will have to be interaction terms, since the effect of 2 inches of snow during rush hour could be different from that in the middle of the day or late at night. Similar methods can also be used to fill in travel times for links which generate no probe reports at all, e.g., some left turn lanes.

9. CONCLUDING REMARKS

In a dynamic route guidance system, the number of real-time information per unit time that is available from the network is limited. Therefore, autonomous route guidance has to be enabled at the level of the vehicle without the intervention of a central monitoring mechanism. In the ADVANCE project, the method employed will be to provide static information, called profiles, in CD-ROMS to the in-vehicle unit. Profiles are estimates of travel times based on the distribution of travel times less the area which falls under the purview of dynamic updates. The reason why this is so and the process by which this will be done is discussed in some detail. The initial CD-ROM updates are output from a network equilibrium model, which will "launch" the process. The method discussed in this paper relates to the updates of static estimates of travel times that will be made to the base network equilibrium estimates once travel time data from roving probes and other sources of real information become available. Finally, operational issues are considered in length to take into account factors that are believed to have ramifications on the accuracy of the estimates.

REFERENCES

1. P. Thakuriah, J. Li and A. Sen, Development of a short-term travel time prediction system for a dynamic route guidance system, Paper presented at the *Regional Science Association International Annual Conference*, Chicago, November, 1992.
2. N. Rouphail, Travel time distributions on signalized links: Applications for ADVANCE, *ADVANCE Working Paper Series*, No. 8, Urban Transportation Center, University of Illinois at Chicago, (1991).
3. J.E. Hicks, D.E. Boyce and A. Sen, Static network equilibrium models and analyses for the design for dynamic route guidance systems, Technical Report in Support of the Design Phase of the ADVANCE Project to the Illinois Department of Transportation, University of Illinois at Chicago, Illinois Universities Transportation Research Consortium, (1992).
4. P.M. Lee, *Bayesian Statistics: An Introduction*, Oxford University Press, New York, (1989).
5. A.K. Sen and M.S. Srivastava, On tests for detecting change in mean, *The Annals of Statistics* **3** (1), 98–108 (1975).
6. A.K. Sen and M.S. Srivastava, Some one-sided tests for change in level, *Technometrics* **17**, 61–64 (1975).
7. A.K. Sen and M.S. Srivastava, On tests for detecting change in mean when variance is unknown, *Annals of the Institute of Statistical Mathematics* **27** (3), 479–486 (1975).
8. H. Chernoff and S. Zacks, Estimating the current mean of a normal distribution which is subjected to change in time, *Annals of Mathematical Statistics* **35**, 999–1018 (1964).
9. G.K. Bhattacharya and R.A. Johnson, Nonparametric tests for shift at unknown time point, *Annals of Mathematical Statistics* **39**, 1731–1743 (1968).
10. A.K. Sen and M.S. Srivastava, *Regression Analysis: Theory, Methods and Applications*, Springer-Verlag, New York, (1990).